Cognition 242 (2024) 105665

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Creating ad hoc graphical representations of number

Sebastian Holt [a,*], Judith E. Fan [a,b], David Barner [a,c]

[a] Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[b] Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, USA
[c] Department of Linguistics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Keywords:
Numerals
Drawing
Communication
Counting
Symbols

ABSTRACT

The ability to communicate about exact number is critical to many modern human practices spanning science, industry, and politics. Although some early numeral systems used 1-to-1 correspondence (e.g., 'IIII' to represent 4), most systems provide compact representations via more arbitrary conventions (e.g., '7' and 'VII'). When people are unable to rely on conventional numerals, however, what strategies do they initially use to communicate number? Across three experiments, participants used pictures to communicate about visual arrays of objects containing 1–16 items, either by producing freehand drawings or combining sets of visual tokens. We analyzed how the pictures they produced varied as a function of communicative need (Experiment 1), spatial regularities in the arrays (Experiment 2), and visual properties of tokens (Experiment 3). In Experiment 1, we found that participants often expressed number in the form of 1-to-1 representations, but sometimes also exploited the configuration of sets. In Experiment 2, this strategy of using configural cues was exaggerated when sets were especially large, and when the cues were predictably correlated with number. Finally, in Experiment 3, participants readily adopted salient numerical features of objects (e.g., four-leaf clover) and generally combined them in a cumulative-additive manner. Taken together, these findings corroborate historical evidence that humans exploit correlates of number in the external environment – such as shape, configural cues, or 1-to-1 correspondence – as the basis for innovating more abstract number representations.

Representations of large exact number - like the numerals *15* and *2028* - have fueled some of the most important achievements of human societies. By allowing humans to represent large quantities exactly, numerals have provided a critical foundation to the development of mathematics and science, the coordination of economic activity in large groups, and the development of human infrastructure and technology. However, humans have not always had the capacity to precisely encode numbers, and many humans currently remain either partially or fully innumerate (Frank, Everett, Fedorenko, & Gibson, 2008; Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004). Further, the anthropological record indicates that written numeral systems evolved slowly over a period of thousands of years, concurrent with changes in complex social institutions and concrete material environments (Beller et al., 2018; Chrisomalis, 2010; Høyrup, 1994; Malafouris, 2010; Robson, 2008; Schmandt-Besserat, 1978, 2010). Finally, experimental research in psychology indicates that, prior to acquiring verbal and written numerals, children are unable to represent exact quantities larger than 3–4, and that in absence of such symbols, perception of number is severely limited (Carey, 2004; Carey & Barner, 2019; Feigenson, Dehaene, & Spelke, 2004). These observations raise the question of how our ancestors converged upon numerals as a solution to representing exact number, and why such innovation was so elusive in the history of our species. What psychological mechanisms might explain the process by which humans create and share novel representations of large exact number?

In the absence of a numeral system, humans, like many other species, represent large quantities approximately, relying on the approximate magnitude system (or AMS). While the AMS is able to represent quantities up to 3 or 4 with precision, representations become noisier as quantities increase, and sets are discriminated on the basis of their ratio, compatible with Weber's law (Dehaene, 1997; Whalen, Gallistel, & Gelman, 1999). Evidence of the AMS is found across human cultures, independent of their adoption of exact numerical symbols. First, when numerate adults are asked to discriminate two sets of rapidly presented dots, they readily identify the larger set when the ratio between the sets is large (e.g., 40 vs. 80 dots, or 2:1), but struggle as the ratio becomes smaller (e.g., 70 vs. 80 dots, or 7:8). Second, innumerate adults, such as the Mundurukú, an Amazonian indigenous group, easily discriminate

* Corresponding author at: Department of Psychology, 9500 Gilman Drive, La Jolla, CA, USA.
E-mail address: sholt@ucsd.edu (S. Holt).

https://doi.org/10.1016/j.cognition.2023.105665
Received 4 July 2023; Received in revised form 9 November 2023; Accepted 11 November 2023
Available online 21 November 2023

dot arrays that stand in a 2:1 ratio (e.g., 20 dots vs. 10 dots), but show declining performance for tighter ratios, like 3:2 (Pica et al., 2004). Similar results have been found in studies of innumerate Pirahã adults. For example, when shown a set of 8 objects and asked to match that set with another set of objects, Pirahã participants often provide approximate matches, and are precise for only the smallest sets up to around 3–4 (Gordon, 2004; Everett & Madora, 2012; cf. Frank et al., 2008). Similar results are found in US children who have not yet learned to count (Schneider, Brockbank, Feiman, & Barner, 2022). Finally, evidence for the AMS is also found in preverbal human infants and in non-human animals, including birds, rodents, and fish (Brannon & Merritt, 2011; Dehaene, 1997).

Numeral systems transcend the limits of the AMS by providing symbols and operations that differentiate large quantities exactly. Verbal numerals, body count systems (Bender & Beller, 2011; Comrie, 2011; Epps, 2006; Hammarström, 2010; Saxe, 1981), written numerals (Changizi & Shimojo, 2005; Chrisomalis, 2020; Ifrah, 2000), and physical calculators like the abacus (Frank and Barner, 2012; Hatano & Osawa, 1983; Stigler, 1984), use similar strategies to extend the human ability to quantify large sets (for reviews, see Barner, 2017; O'Shaughnessy, Gibson, & Piantadosi, 2021). Often, these systems use 1-to-1 correspondence to represent the smallest numbers - using 4 fingers, abacus beads, or vertical lines to represent sets of 4 things (see Fig. 1A). For example, clay envelopes created as early as 11,000 years ago in Mesopotamia often used 1-to-1 correspondence to indicate the number of tokens they contained (Schmandt-Besserat, 2010). Also, a 1-to-1 strategy is found in more recent written number systems such as the familiar Roman numerals (e.g., I, II, III), but also in ancient Greek, Hittite, Cretan, Aramaic, Mayan, and other systems (Ifrah, 2000). When such systems are extended to represent larger numbers, they generally do so in one of two ways. Some systems use configural strategies such as horizontal spacing (e.g. 'IIII II' to represent '6') or stroke directionality (e.g. '~~IIII~~ I'; Chrisomalis, 2020). Also, they generally exhibit similar rules, only allowing new chunks to be created when all previous chunks attain a maximum value (e.g., allowing III III II, but not IIII II II). Other systems however, use arbitrary conventions to express larger numbers. For example, the Roman numeral system represents 5 as V and 10 as X. Likewise, the Soroban abacus (Fig. 1B) uses horizontal space to represent place value, such that some beads have a value of 1, others a value of 10, and others 100, and uses vertical space to assign some beads (i.e., "heavenly beads") values of 5, 50, 500, etc. (e.g., Frank and Barner, 2012; Hatano & Osawa, 1983; Stigler, 1984). Finally, count systems based on the human body typically use arbitrary positions to represent numbers beyond 20 (Bender & Beller, 2011; Saxe, 1981).

Such examples raise the question of how arbitrary conventions arise that transcend 1-to-1 correspondence. According to one cultural-historical account, arbitrary conventions emerge from forms that are initially grounded in contextually available ad hoc comparisons to numerically correlated features of the environment (Cooperrider & Gentner, 2019). For example, measurement terms like "foot" are often created via comparisons to immediately available physical objects (e.g., one's own foot). These concrete measuring conventions are then shared with other people via acts of communication, which leads to standardization (e.g., defining a "foot" as the same length regardless of who uses it), and systematization, where they become defined by other units within a broader network of concepts rather than by their initial, non-arbitrary, ad hoc comparison (e.g., a foot is defined as 12 in., $^1/_3$ of a yard, etc., but *not* the foot of whoever is making the measurement). Number words often follow a similar trajectory, beginning via ad hoc comparison to concrete objects, but culminating in more arbitrary conventions over time as they become integrated to a new system of use. For example, in the Hup language (spoken in Colombia and Brazil) the word for 1 originated from the demonstrative term "that", 2 from "eye-quantity", and 3 from "rubber tree seed quantity" (since seed pods contain 3 seeds; Epps, 2006). Similarly, 10 is expressed as "both hands" and 20 as "both feet" - a widely attested practice in the cultural history of number (see also Bender & Beller, 2012; O'Shaughnessy et al., 2021; Saxe & Esmonde, 2012; Williams, 1940). However, as in the case of measure words, these expressions that began as ad hoc comparisons have subsequently changed, becoming phonologically distinct from the words that served as their historical basis, and used strictly for the abstract function of referring to number (and not to eyes, seeds, etc.). Such facts suggest that ad hoc reference is a critical first step in the process of creating number conventions, followed by the derivation of abstract meanings from their position within the system of symbols (Damerow, 1996).

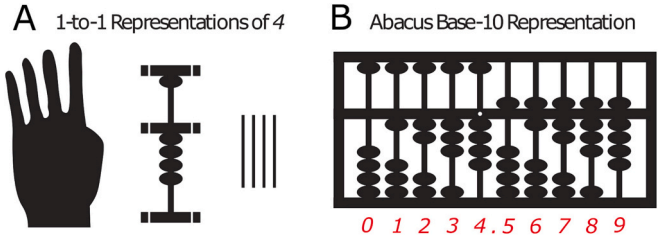

Fig. 1. A. Three 1-to-1 modes of representing the quantity 4: fingers, abacus beads, and written strokes. B. A soroban showing different digits in base-10. Much like the western Arabic numerals, each column denotes a place-value, such that the column showing *3* represents a magnitude of *30* while the column showing *4* represents a magnitude *4*. The white dot indicates the ones-place, to the right of which columns represent decimal values.

While the emergence of numerical conventions in cultural history is a topic of significant debate among anthropologists and historians (see Chrisomalis, 2004, for review), it is less frequently studied by cognitive psychologists, and we know of no systematic experimental study of how humans create novel numerals. Psychologists frequently discuss the evolutionary origins of numerical perception (Gallistel, Gelman, & Cordes, 2006), but few discuss the role of cognitive factors in the cultural evolution of numeral systems, with some notable exceptions (for some discussion of this topic, see Barner, 2017; Hurford, 1987; O'Shaughnessy et al., 2021; Xu, Liu, & Regier, 2020). For example, in one recent study, Xu et al. (2020) analyzed the verbal numeral systems of 30 currently spoken languages, and argued that the evolution of cross-linguistic variation is shaped by a need for precise yet cognitively efficient communication. Similarly, O'Shaughnessy et al. investigated the cultural origins of symbolic number by characterizing the variability in attested numeral systems, and reviewing known differences in how different cultures and linguistic forms express large and small numbers. However, while experimental methods have sometimes been used to characterize variability in existing numeral systems (e.g., Bender & Beller, 2012; Frank et al., 2008; Gibson et al., 2019; Gordon, 2004; Pica et al., 2004; Saxe & Esmonde, 2012), no previous work has investigated the creation of novel numerals experimentally.

Although there is little experimental work related to the evolution of numerical symbols, a larger literature has used experimental methods to investigate the origin of symbols representing non-numerical concepts (Christiansen & Kirby, 2003; Kirby, Cornish, & Smith, 2008). Multiple studies have found that abstract conventional symbols emerge gradually in interactive communicative games. In Pictionary-style games, while participants initially create drawings that are transparently related to the things they represent, later drawings tend to become more abstract and conventionalized (Fay, Garrod, Roberts, & Swoboda, 2010; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007; Hawkins, Sano, Goodman, & Fan, 2023). An initial reliance on non-arbitrary symbols can also be found in novel communicative modalities, with which participants have no previous experience. For example, participants have been found to use more complex auditory signals to refer to more complex objects (Hofer & Levy, 2019; Verhoef, Kirby, & De Boer, 2016).

Here we adapt these experimental methods to investigate how participants use a graphical communication medium (i.e., involving mark-making and/or pictures) to convey number. In particular, we asked how

participants express number when prevented from using pre-existing conventions (e.g., Arabic numerals), and how their ability to communicate is impacted by the availability of physical shapes and configurations that could be used as proxies for number, akin to the use of hands, feet, or words for objects like eyes and seeds. In Experiment 1, participants were paired up to play a drawing-based communication game in which one participant (the sketcher) sought to communicate about an array of objects to another participant who could not see the array (the viewer). In "number games," only the number of objects was relevant; in "shape games," only the identity of the objects was relevant. We found that drawings produced in the number games were markedly different from those in the shape games: they were generally composed of marks that stood in 1-to-1 correspondence with the objects but no longer resembled them - a form of abstraction that is typical in attested numeral systems. Moreover, some sketches seemed to mirror the spatial configuration of target sets, while others arranged the objects into orderly configurations. Following up on this observation, Experiment 2 further explored the conditions under which participants might exploit configural cues to express number. When the sketcher was shown target arrays in the same configuration as the viewer, and was aware of this fact, sketchers frequently exploited these configural cues to use strategies other than 1-to-1 correspondence. While Experiment 2 provided an in-principle demonstration that participants used shared features of their physical environment to communicate number, it did so in a way that is likely rare in the wild, as most sets seldom occur in reliable configurations like rows and columns. Also, in both Experiments 1 and 2 we required participants to create representations of number de novo, whereas people often use existing objects and their labels to express number via ad hoc comparison to other objects. In keeping with these considerations, in Experiment 3 we asked participants to communicate number using a menu of familiar shapes that differed according to numerically salient features (e.g., a pair of cherries, a four leaf clover, etc.). We measured to what degree participants would seize upon these numerical features to transcend the use of 1-to-1 correspondence, and also what rules of composition they might use to express larger numbers.

1. Experiment 1

In Experiment 1, we conducted a preliminary investigation of how people use ad hoc graphical representations to communicate number (i.e., without the aid of existing symbols). Participants were paired up online to play a sketching-based reference game and, on each trial, were presented with four arrays of objects. One participant (sketcher) aimed to produce a sketch that would help their partner (viewer) identify which array they intended to refer to. To explore the impact of having the specific need to communicate about number information, we manipulated whether these arrays differed in the number or the kind of objects they contained. We sought to characterize what strategies dyads used to communicate when only number information was relevant, by comparison to the baseline scenario in which only object identity was relevant. Towards this end, we both analyzed the visual properties of these sketches and recruited naive participants (recognizer) to interpret them out of context, providing complementary insight into their content and organization.

1.1. Methods

Experimental methods for this and subsequent experiments were pre-registered using the Open Science Framework, available at: https://osf.io/4q3t9

1.1.1. Participants

Participants were recruited for two tasks: a communication task and a recognition task. For the communication task, we initially recruited 134 participants from Amazon Mechanical Turk (AMT), who were paired up to form 67 dyads who interacted with one another throughout an entire experimental session. Data from six sessions were excluded according to our pre-registered criteria: four did not meet the performance threshold of 50% accuracy, while two others did not follow task instructions (e.g., having "drawn" text). Thus data from 122 participants ($N = 61$ sessions) were included in further analysis. Of these participants, 117 individuals completed our optional demographic survey (53 female; age data was not collected due to a technical error). For the recognition task, we recruited a total of 211 participants from Amazon Mechanical Turk. Thirteen recognition participants were excluded according to pre-registered criterion of missing any one of four catch trials, and eight further participants were excluded for having participated in the communication task, leaving 190 participants, of whom 165 completed our optional demographic survey (62 female; $M_{age} = 37.65$ years, $SD_{age} = 10.80$ years). All participants provided informed consent in accordance with our IRB.

1.1.2. Stimuli

Stimuli were visual arrays of objects, each containing a variable number of identical shapes arranged in an arbitrary configuration. These arrays could contain between 1 and 8 items and contain one of 4 different shapes (i.e., bear, deer, owl, rabbit silhouettes), resulting in 32 number × shape combinations. This range of numbers was chosen to include values both above and below the subitizing range of ~3–4. To preclude reliance on spatial cues, the configuration of objects within each array was independently randomized for each participant and across dyads.

1.1.3. Procedure

1.1.3.1. Communication task. The communication task was a sketching-based reference game for two players (Fig. 2A), in which one participant (the *sketcher*) created sketches to help their partner (the *viewer*) identify one 'target' visual array at a time from three similar distractors. While the complete set of target arrays was identical across all games (i.e., every game featured sets of 1–8 bears, deer, owls, or rabbits), the distinguishing feature between target arrays and distractors within each trial varied between two conditions. Some dyads were randomly assigned to the Shape condition ($N = 29$), in which all four visual arrays on a given trial featured the same number of animals but differed in the kind of animal they contained. Other dyads were assigned to the Number condition ($N = 32$), in which all four visual arrays featured the same kind of animal but differed in the number of animals per array. For each array shown to the sketcher, a corresponding array with the same number and type of animal was shown to the viewer, but the spatial configuration of animals was independently randomized for each participant, to preclude reliance on configural cues. Both participants in the dyad therefore had the same basis for inferring the which feature of the visual arrays – number or shape – was relevant.

On each trial, the sketcher used a 500 × 500px digital canvas embedded in a web browser to produce their sketch. They were required to complete this sketch in less than 30 s, after which additional strokes could not be registered. Also, they were told that they should not use existing numerical symbols (e.g., 5, 6, 7) but that they could create new symbols of their own (for complete instructions, see Supplemental Materials, section S1.1). After the sketch was complete, it was shown to the viewer, who was then asked to select the array they thought was the target, without a time limit. Then both participants received feedback: the viewer was shown the identity of the target and the sketcher was shown which array the viewer selected. There were 32 trials in each game, such that each of the 32 number × shape combinations served as the target exactly once. These 32 trials were divided into four blocks of eight trials each, such that each number appeared once within each block and each animal appeared twice.

1.1.3.2. Recognition task. Participants in the recognition task were

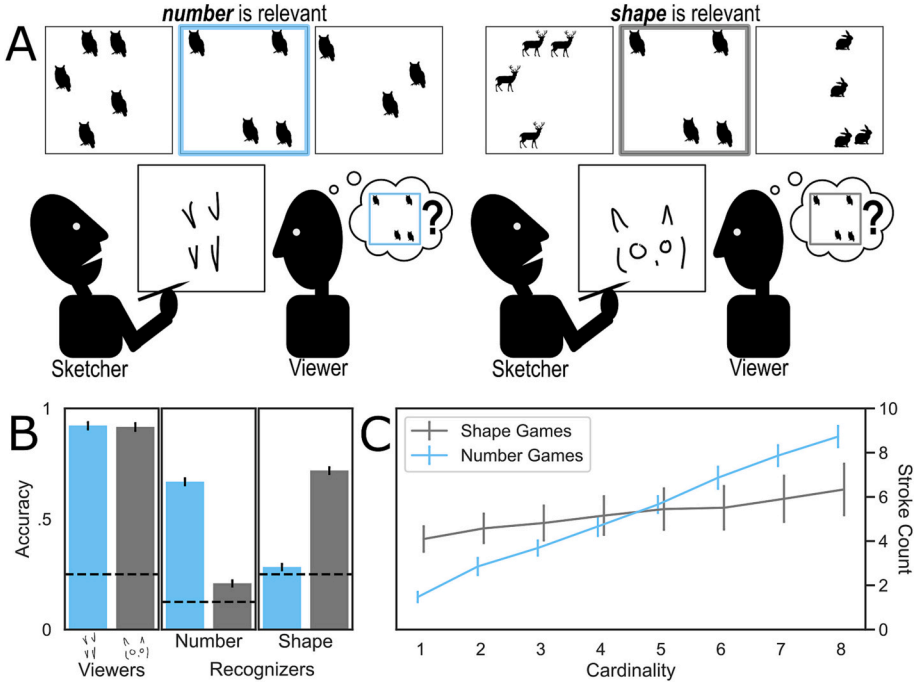

Fig. 2. A. In the number condition (left), sketchers had to communicate one target image to a viewer from a set of images that differed in the number of objects. In the shape condition (right), the set of images differed in the kind of object, but not the number. B. Left panel: viewer accuracy for games in each communication task condition. Middle panel: Recognizer accuracy in providing number labels for sketches produced in the number condition (blue) and shape condition (gray). Chance performance in dotted lines. Right panel: same as middle, but for recognizers providing shape labels. C. Average number of strokes in sketches for each tested cardinality in both number (blue) and shape (gray) games. While a weak correlation exists in shape games, it is almost exactly 1 in number games (error bars are 95% CI in all panels). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presented with sketches produced in the communication task, and were asked to identify shape or number information in each sketch. On each trial, these participants were presented with a single sketch and several buttons below. Buttons were either labeled with Arabic numerals or animal shapes, depending on whether the participant had been assigned to the Number or Shape recognition condition. Participants in the Number group judged which numeral lying in the range 1–8 best matched each sketch; participants in the Shape group judged which shape (i.e., bear, deer, owl, or rabbit) provided the best match.

Each participant was presented with a total of 61 sketches, one randomly sampled from each communication game (N = 61 dyads). They were thus presented with sketches from both Number and Shape communication games, and had to guess intended meanings while naïve to the communicative context in which the sketches were originally produced. As an attention check, four catch trials were inserted at regular intervals throughout each session, where participants were presented with a sketch that matched one of the buttons exactly. Participants were included in analysis only if they selected the matching button in all four of these catch trials.

1.2. Results

1.2.1. Communicative strategy

We first asked whether participants communicated number or shape information comparably. To this end, we measured (1) how accurate pairs were, (2) how long sketchers took, (3) how much virtual 'ink' sketchers used, and (4) how many strokes sketchers used. We constructed mixed-effects regression models to predict each of the above outcome variables, based on: communication condition (Number vs. Shape), trial block number (i.e., 1–4), shape (i.e., bear, deer, owl, rabbit), and the number of objects in the target array (i.e., 1–8). In all linear models, variation across games was modeled by fitting a random intercept for each game.

These models revealed similarly high accuracy in both conditions (b = −0.221, z = −0.623, p = .534; see Fig. 2B), but that Shape game sketchers used more time and 'ink' to create their sketches (time: b = 6.288, t = 7.675, p < .001; ink: b = 0.034, t = 8.98, p < .001). Over successive trial blocks, dyads in both conditions became more accurate (b = 0.494, z = 5.745, p < .001), took less time (b = −1.369, t = −23.051, p < .001), used less ink (b = −0.001, t = −4.742, p < .001), and made fewer strokes (b = −0.553, t = −11.029, p < .001). These results suggest that communication of number and shape was functionally comparable, both in getting the point across and in doing so better over time.

We found that participants were more accurate for smaller cardinalities (b = −0.252, z = −6.001, p < .001), and produced more strokes for larger cardinalities in Number games (b = 1.029, t = 15.782, p < .001) but not Shape games (b = − 0.734, t = −7.767, p < .001), suggesting that a 1-to-1 strategy was used only for communicating number (Fig. 2C). To explore this, we therefore analyzed the 'stroke ratio' of sketches. We reasoned that if the ratio of strokes (in a sketch) to objects (in a target) was 1, this would indicate the use of 1-to-1 correspondence (e.g., IIII to represent 4 objects), while a smaller number of strokes would indicate the use of a compressed form (e.g., IV). With this measure, we found that 77.2% of Number-game sketches used 1-to-1 correspondence (CI: [74.7, 79.8]) versus only 14.3% of Shape-game sketches (CI: [12.1, 16.6]).

A qualitative analysis of the sketches (see Fig. 3) produced in these games revealed that Number-game sketches sometimes used strategies that we had not anticipated. One strategy featured the use of 1-to-1 correspondence, but was partially depictive, such that the configuration of objects within the target array was mirrored by the configuration of strokes within the sketch. Another strategy preserved 1-to-1 correspondence, but not configural information, instead rearranging strokes into patterns that appeared more geometrically regular than the random arrays of objects in the target. A rarer strategy appeared to make use of geometric shapes that correlated with the cardinality of the target array. For example, one participant used a five-pointed star to communicate an

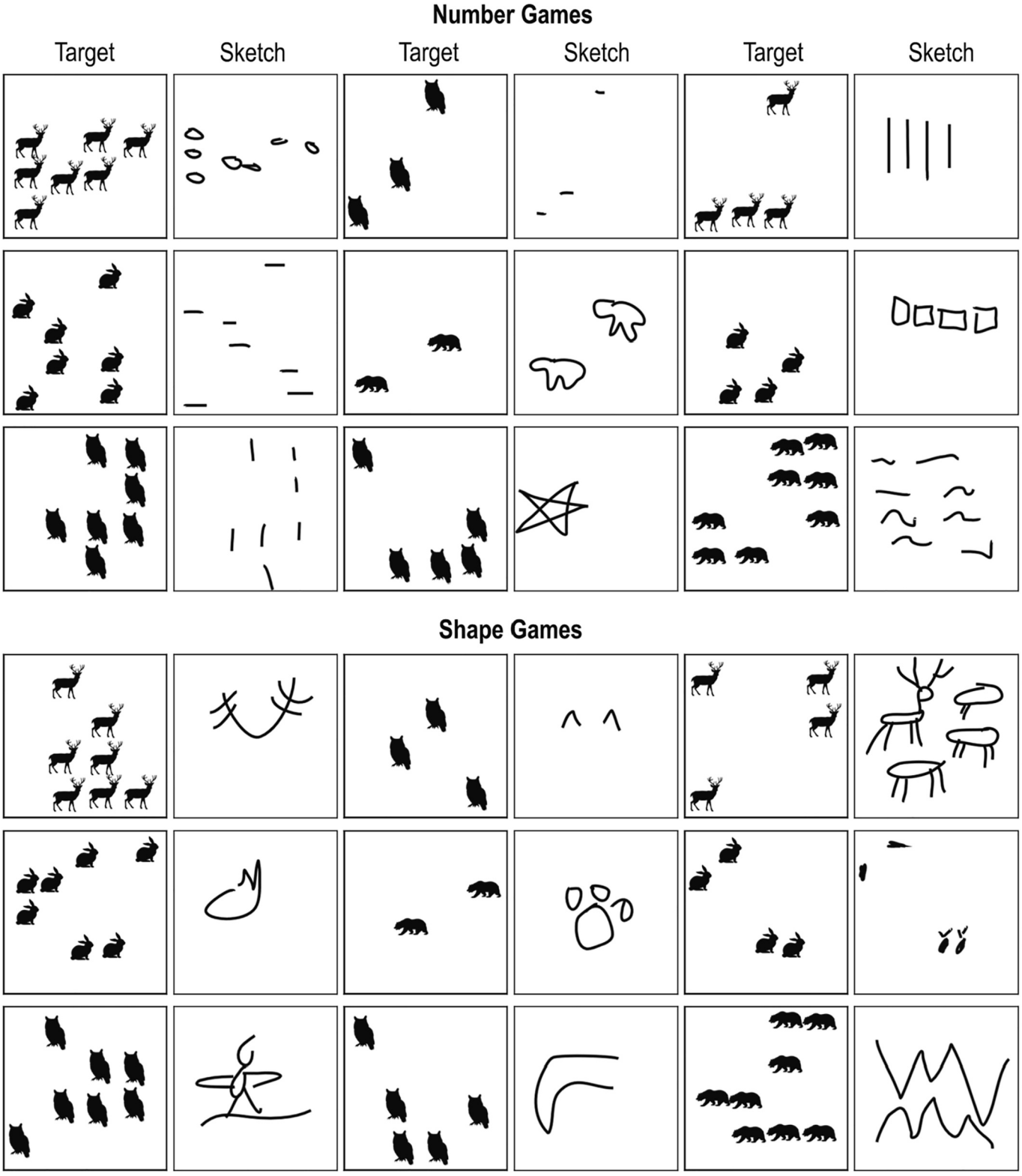

Fig. 3. Some Example Sketches. Above. A selection of nine target arrays from number game trials, with the sketch made for each to its right. Below. Same as above, but shape game trials.

array of 5 owls. In general, such strategies are consistent with historical observations that humans draw on publicly available correlations in the environment to communicate about number (e.g., Epps, 2006).

1.2.2. Message content

We next asked how much number or shape information was included in sketches, and whether this was affected by the relevance of each type of information (e.g., did Shape-game sketchers include irrelevant numerical information?). To answer this question, we asked each recognizer in the recognition task to recover one type of information – number or shape – from sketches produced in the communication task. Each recognizer was presented with one sketch from every game (both

Number and Shape games), and we measured the content of sketches by recognizers' ability to recover this information.

We constructed a binomial linear mixed-effects model, which included communication game condition (i.e., number or shape), recognition task condition (i.e., number or shape), communication game trial block number (i.e., 1–4), and the interaction between communication game condition and recognition task condition as fixed-effects predictors, and modeled variation between individual recognizers, communication games, stimulus cardinality, and shape using random intercepts.

Overall recognition accuracy was slightly lower for shape games than for number games ($b = -2.364$, $z = -23.860$, $p < .001$), and was also lower for recognizers trying to recover shape information rather than number information ($b = -1.732$, $z = -20.109$, $p < .001$). Crucially, recognizers in both conditions were better at recovering information that was relevant in the original communication task than information that was not relevant ($b = 4.362$, $z = 50.604$, $p < .001$; Fig. 2B), suggesting that Number-game sketches included much more number information than shape information (and vice-versa). In fact, recovery of irrelevant information was close to chance (0.25) in shape games at 20.8% (CI: [19.4, 22.3]), and also low, although significantly higher than chance (0.125), in number games at 28.2% (CI: [26.8, 29.7]). Finally, this model also revealed that recognizers were better able to recover information from sketches produced early in communication games than sketches produced towards the end of the game ($b = -0.070$, $z = -3.819$, $p < .001$), which suggests the emergence of some dyad-specific conventions over the course of the communication games.

1.3. Discussion

In this experiment, we found that participants communicated effectively about both number and shape, when each feature was communicatively relevant. To communicate about number, they overwhelmingly relied on 1-to-1 correspondence, ignoring the shapes of individual objects (and vice versa for shape). Also, we found that larger numbers were communicated less accurately than smaller ones. This is consistent with prior work that has found that larger cardinalities are associated with noisier magnitude representations, particularly under speeded conditions when counting is impractical. Finally, some dyads generated novel strategies for encoding number and shape, as evidenced by the fact that the representations they used were not interpretable to independent "recognizers". For example, some participants appear to have taken advantage of the correlation between spatial configurations and cardinality in order to create depictive, configural representations of number.

These results are interesting for a number of reasons. First, despite the fact that all of the participants in our studies were fully numerate adults who were familiar with at least one numerical notation system (i.e., the Arabic numerals) it was clearly not trivial to create arbitrary representations of number de novo. This result provides experimental evidence that creating novel conventions for number is a qualitatively harder task than creating conventions for representing familiar objects. It also establishes that this method can be used to explore how participants might overcome this challenge. Second, the results are interesting because they demonstrate that the strategies preferred by participants align with the most common early strategies used by human ancestors - i.e., 1-to-1 correspondence that abstracts away from the identity of the items being enumerated, and using correlations in the environment as a proxy for cardinality. Participants created arrays of dots that ignored shape information almost universally, similar to systems like those attested in Greek, Hittite, Cretan, Aramaic, Mayan, etc. Also, although 1-to-1 was readily available and easy to deploy, some dyads nevertheless innovated, used configural cues, and produced forms that could not be decoded by others, potentially compatible with the creation of novel conventions.

In Experiment 2, we explored this last finding, and began to probe how participants might make use of publicly shared correlates of number to transcend the reliance on 1-to-1 correspondence. The history of numerical symbols indicates that 1-to-1 representations sometimes evolve over time into more arbitrary conventions by exploiting visual regularities (Chrisomalis, 2020). For example, the way marks are consistently arranged on each face of a 6-sided die allows for rapid and efficient decoding and communication of numerical information. Such configural cues are also exploited in the design of the abacus to facilitate the recognition and manipulation of larger cardinalities (Frank & Barner, 2012; Srinivasan, Wagner, Frank, & Barner, 2018), and in some early writing systems, as described in the Introduction. In each case, rather than counting each mark, people can map a recognizable visual pattern onto a specific cardinality. Because humans are adept at exploiting incidental correlations to communicate quantity (Cooperrider & Gentner, 2019), it is possible that the presence of visual regularities of sets in the world could facilitate the creation of compressed representations that exploit them. We explored this in Experiment 2 by creating contexts that enhanced the availability of configural cues, while also increasing the incentive to innovate by presenting participants with larger sets.

2. Experiment 2

In Experiment 2, we manipulated the visual regularity of sets to ask whether this might lead participants to produce sketches that were more compressed (i.e., using fewer strokes to describe the same number of objects), rather than relying on 1-to-1 correspondence. Specifically, we manipulated whether sets of animals were arranged in a regular grid pattern with columns of 6 animals each, or were instead arranged in an irregular configuration, similar to Experiment 1. We reasoned that regular configurations might promote compression, e.g., by allowing participants to replace complete columns of dots with vertical lines, or rows with horizontal lines, or to use shapes (like rectangles) to indicate which grid of dots is intended. In addition, we shifted the range of cardinalities that were tested from 1 to 8 (in Experiment 1) to sets that were as large as 20 (in Experiment 2). We reasoned that, given that 1-to-1 is a common strategy for representing smaller numbers, participants might only have incentive to create compressed representations when the cost of 1-to-1 grows, as with larger numbers. Moreover, because people in serial reproduction experiments implicitly introduce regularity into symbolic systems that render those systems easier for others to learn (e.g., Kirby et al., 2008; see Wang, Lew, Brady, & Vul, 2023 who show this with dot arrays similar to our stimuli), we expected stimuli that already feature visual regularity to be easier for sketchers not only to communicate about, but also to propose compressed forms that viewers could accurately decode.

2.1. Methods

2.1.1. Participants

We recruited 200 participants on AMT, as in Experiment 1. Of the resulting 100 games, 41 were excluded per our pre-registered criteria: 35 for scoring below our accuracy threshold of 50% (same as Experiment 1),[1] and 6 for using existing numerical symbols. This left 29 irregular condition games and 30 regular condition games. Of the remaining participants, 113 answered the optional demographic survey (50 female; $M_{age} = 36.46$ years, $SD_{age} = 11.07$ years). Finally, an error in recording reaction times affected 3 of the remaining games. Note that while analyses reported below reflect these exclusions, results did not

[1] This somewhat high exclusion rate might be explained by some participants missing the part of the task instructions which explained that the positions of the arrays were independently randomized for the sketcher and viewer. For example, some sketchers always drew 1–4 marks, corresponding with the position of the target array among the distractors

differ when data from these last 3 games were included.

2.1.2. Design, materials, and procedure

2.1.2.1. Materials. Stimuli were generated as in Experiment 1, with three differences. First, only larger cardinalities (15–20) were used, rather than smaller ones (i.e., 1–8). Second, these cardinalities were presented in two different ways: For games in the Irregular condition, the spatial layout of animal silhouettes was again generated by sampling locations for each item at random as in Experiment 1. For games in the Regular condition, however, the animal silhouettes were arranged in a rectangular grid pattern, with each column containing 6 items and the rightmost column containing the remainder (e.g., 15 had a remainder of 3 animals in the rightmost column; Fig. 4A). Finally, only three of the previous animal types were used (i.e., bear, deer, and owl, but not rabbit).

2.1.2.2. Procedure. The procedure was identical to Experiment 1, with two important differences. First, participants saw each animal-cardinality combination twice, once in the first half of the experiment (18 trials) and once in the second half (18 more trials), rather than only once as in Experiment 1. Further, each block of 6 trials contained one instance of every cardinality as the target. Apart from these constraints, the presentation order of stimuli was randomly shuffled within each block of 18 trials. Second, sketchers and viewers saw each array in the same spatial configuration as each other, allowing configural information to be leveraged for communication. This was impossible in Experiment 1, as spatial configuration was randomized independently for each partner. Furthermore, for each sketcher-viewer dyad, arrays of each cardinality were consistently presented in the same spatial configuration across trials, making it possible for sketchers to use the same configural cues to communicate about particular cardinalities across trials – and thus to develop ad hoc graphical conventions. We did not deem it necessary to explicitly inform participants of this fact, based on our observation that in Experiment 1 most participants in the number condition produced drawings that preserved information about how objects were arranged in the target array, even when this information was irrelevant (because the spatial configuration of objects shown to the viewer was different).

2.2. Results[2]

2.2.1. Communicative efficiency

We first investigated communicative efficiency, which we measured in three ways: (1) how accurate pairs were, (2) how long sketchers took to make their sketch, and (3) how long viewers took to submit their guess. We found that dyads in both conditions achieved high accuracy (Regular: 83.7%, CI: [81.5, 85.9]; Irregular: 87.8%, CI: [85.9, 89.8]). We constructed a binomial linear mixed-effects model to predict communicative success similar to Experiment 1, but also included an interaction term between regularity and cardinality. As in Experiment 1, this model revealed that dyads were more accurate when communicating small numbers than large numbers ($b = -0.122$, $z = -2.020$, $p = .043$), more accurate in later trial blocks ($b = 0.421$, $z = 9.760$, $p < .001$), and more accurate in the Irregular condition than in the Regular condition ($b = -4.604$, $z = -3.173$, $p = .002$). Sketchers took less time to communicate regular arrays than irregular arrays ($b = -4.363$, $t = -3.23$, $p = .002$), although there was no difference between conditions in the time that viewers took to make their guesses ($b = 22.1036$, $t = 1.349$, $p = .178$). This result could reflect the fact that sketchers could estimate arrays more quickly when organized into groups (Atkinson, Campbell, & Francis, 1976; Ciccione & Dehaene, 2020; Mandler & Shebo, 1982; Starkey & McCandliss, 2014; Van Oeffelen & Vos, 1982) or that sketches of regular arrays can be drawn more quickly than sketches of irregular arrays. Together with the model predicting accuracy, these results suggest that sketchers presented with regular visual arrays were quicker than those shown irregular arrays, but that they communicated number information less reliably (Fig. 4B & 4C). Additionally, while viewers were slower and less accurate as target arrays grew in size, this effect was smaller in the Regular condition than in the Irregular condition (response time:$b = -1.709$, $t = -2.295$, $p = .022$; accuracy: $b = 0.236$, $z = 2.942$, $p = .003$). These results suggest that regular arrangements of objects provide a more scalable representation of large numbers, allowing viewers to decode the meaning of visual representations faster despite the increasing size of sets and complexity of representations.

2.2.2. Communicative strategy

We next investigated the degree to which participants used 1-to-1 or more compressed strategies to communicate about number, such as relying on configural cues. As in Experiment 1, we did so by computing stroke ratios, which should be close to 1 when participants use a 1-to-1 strategy and smaller than 1 when they use compressed strategies. Unlike the Number condition of Experiment 1, in which participants used a 1-to-1 strategy to communicate about number in 77.2% of trials, participants in Experiment 2 used much less frequently in both the Irregular condition (17.5%, CI: [15.2, 19.8]) and the Regular condition (16.8%, CI: [14.5, 19.0]). Instead, compressed strategies were common: whereas Number condition participants in Experiment 1 employed compression in only 2.9% of sketches, participants in Experiment 2 employed compression frequently in both the Irregular (77.0%, CI: [74.5, 79.6]) and Regular (78.2%, CI: [75.7, 80.6]) conditions.

We constructed a linear mixed-effect model to predict the stroke ratio of each sketch, which revealed that participants employed compression to a similar degree regardless of the visual regularity of arrays they had to communicate ($b = -0.065$, $t = -0.993$, $p = .325$). Also, representations of larger cardinalities were significantly more compressed than representations of smaller ones ($b = -0.014$, $t = -6.760$, $p < .001$), and compression was greater in later trial blocks ($b = -0.013$, $t = -6.393$, $p < .001$). Together, these results suggest that as participants gained experience with the task, they became more effective at encoding number in their sketches, and that the regularity of spatial arrangements had no effect on either the choice of strategies or the amount of compression employed when participants did depart from a 1-to-1 strategy.

2.3. Discussion

Experiment 2 asked whether visual regularities in an array of objects help participants create compressed representations of number. While participants in Experiment 1 were reluctant to communicate about number information using strategies other than 1-to-1 correspondence, sketches in Experiment 2 were much more compressed, relying substantially less on 1-to-1 correspondence. Two features of Experiment 2 may explain this finding. First, arrays in Experiment 2 were presented in configurations - either regular or irregular - that were shared between the sketcher and the viewer. Second, the sets in Experiment 2 were larger than those in Experiment 1. Compatible with the idea that this may have promoted compression, participants in Experiment 2 were significantly more likely to use compressed representations for larger sets than for smaller ones. This suggests that the incentive to create compressed representations of number may be greatest when individuals are asked to communicate about large quantities - a fact that is compatible with the historical record, in which systems often used 1-to-1 strategies for small numbers, but compressed conventions for larger ones.

In both Experiments 1 and 2, we provided participants with

[2] Two pre-registered analyses are not reported. The first was a spatial clustering analysis to measure visual chunking, and the second was a classification task asking online participants to infer sketch strategies, as in Experiment 1; both proved to be uninterpretable.

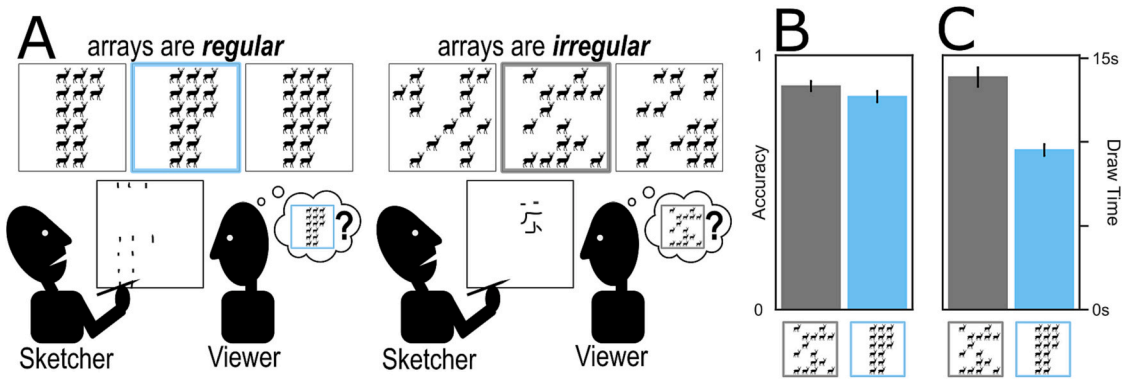

Fig. 4. A. In the regular condition (left), all arrays were organized into columns of 6, with a remainder column on the right. In the irregular condition (right), arrays were organized randomly, as in Experiment 1. B. Viewers were slightly less accurate when decoding sketches of regular arrays (error bars are 95% CI). C. Sketchers made sketches in much less time when encoding regular arrays than irregular ones (error bars are 95% CI).

numerical arrays and asked them to spontaneously generate novel symbols that might be used to represent number. These studies found evidence that participants are eager to use shared correlates of number to communicate, but also that generating conventions to represent number de novo is no easy task in such contexts. However, as noted in the Introduction, evidence from the historical record indicates that conventions for labeling number and numerical measures rarely emerged from purely arbitrary innovations, and very frequently arose by processes of ad hoc comparison to ubiquitous, publicly available objects, like body parts. Also, although numeral systems sometimes exploit configural cues to represent number, most systems do not, perhaps because configural cues are not reliable properties of the things that humans wish to communicate about.

In Experiment 3, rather than burdening participants with the need to create entirely novel representations from scratch, we probed their interest in exploiting publicly shared correlates of number by providing them with objects that could act as candidate conventions, by virtue of their numerically relevant shapes (e.g., a pair of cherries, or a four-leaf clover). This allowed us to not only explore this idea, but also to test whether, when using such objects to communicate, participants show evidence of using combinatorial rules akin to those found in numeral systems in the attested historical record.

3. Experiment 3

Experiment 3 asked how, if at all, communicators would exploit numerical features of familiar objects in their shared environment to express number, and whether their use of such features might facilitate the use of combinatorial rules. The use of ad hoc comparison to convey quantity is frequently attested in human languages (Cooperrider & Gentner, 2019; Epps, 2006), as is the combination of small number words to express larger quantities (Comrie, 2011). To facilitate the adoption of such ad hoc comparisons, we allowed participants to communicate using visual tokens that featured conspicuous numerical properties (e.g., a pair of cherries, a four leaf clover, etc.). These visual tokens were included on a keyboard that participants used to communicate with a partner, and permitted the combination of multiple, concrete images to express larger numbers. We asked whether participants would persist in using a 1-to-1 correspondence strategy (e.g., 1 clover or pair of cherries per item), or instead use the numerical information available in these images as symbols for number. Also, we asked whether participants would adopt rules for combining these shapes, and if so, which rules.

3.1. Methods

3.1.1. Participants

We recruited 110 participants from Prolific. Twenty-nine games were excluded according to pre-registered criteria of finishing all trials (10 games) and having a minimum of 50% accuracy (19 games), leaving 62 participants paired in 31 games (27 female; $M_{age} = 26.48$ years, $SD_{age} = 8.16$ years). All participants provided informed consent as per the IRB.

3.1.2. Design, materials, and procedure

3.1.2.1. Materials. Stimuli were generated according to the same process as in Experiments 1 and 2, but included both small and large cardinalities (1–16), and contained only dots instead of different kinds of animals. This was because based on the results of Experiments 1 and 2, the exact animals appeared to no longer be theoretically relevant. Rather than a sketchpad, participants were presented with a virtual keyboard consisting of four buttons. Each button featured a different natural object, each of which had a different number of conspicuous physical features ranging from 1 to 4 (Fig. 5A). These were an apple, a pair of cherries, three connected oak leaves, and a four-leaf clover. There was also a 'delete' key. While no explicit limit was told to participants, the task interface limited messages to no more than 30 tokens – a limit which was never reached.

3.1.2.2. Procedure. The procedure was similar to Experiments 1 and 2, in that a *sender* (analogous to the *sketcher*) sent messages to a *receiver* (analogous to the *viewer*) to communicate one image out of a larger set of images. However, it was different in two respects. First, all 16 cardinalities appeared once in the first half of the experiment, and once again in the second half, for a total of 32 trials. Second, we reduced the total number of images presented on each trial from 4 to 3 (i.e., 1 target +2 distractors). This was to discourage participants from trying to communicate the ordinal position of the target image (e.g., 2nd from the left) by using the ordinal position of the four response keys, which were also arranged horizontally. Such a strategy would result in chance performance, as the ordinal position of images was randomized independently between partners. Like Experiment 1, the spatial configuration of dots within each array was also independently randomized for each partner, to preclude reliance on configural cues, as well as between every two occurrences of that array within the task.

3.2. Results

3.2.1. Communicative efficiency

We measured communicative efficiency in three ways: (1) how accurate pairs were, (2) how long senders took to make their message, and

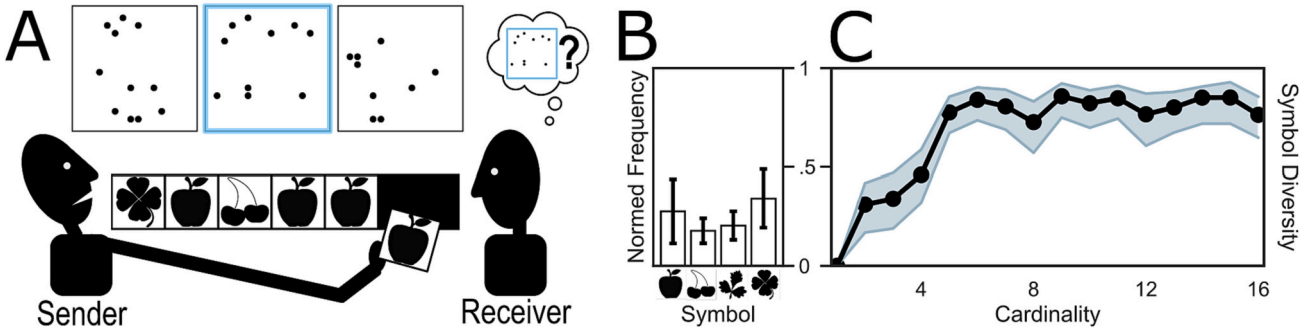

Fig. 5. A. In Experiment 3, participants communicated about arrays of dots that varied in quantity. Rather than a sketchpad, they communicated using a keyboard of existing shapes. B. Relative frequency of each of the 4 shapes across all messages (error bars are 95% CI). C. Simpson's diversity index of messages created to represent sets of each cardinality.

(3) how long receivers took to submit their guess. To do this, we constructed linear mixed effects regressions, modeling the effects of cardinality and trial number on each measured variable. Overall accuracy was slightly lower than in Experiments 1 (91.9%) and 2 (85.7%), at 78.7%. As in Experiments 1 and 2, participants improved at the task as they gained experience with it: dyads became more accurate over successive trials ($b = 0.062$, $z = 6.154$, $p < .001$), senders took less time to compose their messages ($b = -0.180$, $t = -3.119$, $p = .002$), and receivers took less time to make their guesses ($b = -0.207$, $t = -2.374$, $p = .018$). Also similar to Experiments 1 and 2, larger cardinalities were communicated less accurately than smaller ones ($b = -0.147$, $z = -7.13$, $p < .001$), and more slowly both by senders ($b = 1.279$, $t = 11.3$, $p < .001$) and receivers ($b = 1.000$, $t = 5.849$, $p < .001$).

3.2.2. Communicative strategy

We next investigated the degree to which participants used 1-to-1 or compressed strategies to communicate number. Analogous to the stroke ratio measure employed in Experiments 1 and 2, we measured *token ratio*: the ratio between the number of tokens used within a message to the cardinality of its target array. A ratio close to 1 would suggest the use of 1-to-1 correspondence, while a ratio less than 1 would suggest the use of a more compressed strategy. By this measure, participants in Experiment 3 used 1-to-1 strategies in 14.9% of trials, similar to the 17.1% of trials measured in Experiment 2, but much less than the 77.2% of trials in the Number condition of Experiment 1. Removing trials in which the target cardinality was 1 (1/16 of all trials), where the use of 1-to-1 would thus be indistinguishable from other strategies, this proportion was only 9.2% - significantly lower than in Experiment 2 (Fisher's exact test; $p < .001$). Instead, compressed strategies were evident in 84.6% of trials, suggesting that the use of 1-to-1 correspondence was relatively infrequent overall and limited to a small number of games. Finally, senders were also significantly more likely to compress information when they were communicating larger cardinalities than smaller ones ($b = -0.032$, $t = -7.286$, $p < .001$), and this did not change over successive trials ($b = 0.001$, $t = 0.540$, $p = .589$). This distribution of strategies suggests that the presence of objects that support ad hoc comparison helps communicators to employ strategies that bypass direct 1-to-1 correspondence between tokens and objects.

An exploratory linear model predicting accuracy from token ratio revealed that compressed forms did not significantly contribute to higher accuracy ($b = -0.428$, $z = -1.175$, $p = .240$), and may have reduced accuracy as numbers became larger ($b = 0.195$, $z = 3.166$, $p = .0015$). This may be because larger numbers could be expressed in a greater number of ways, as they appear to have been in senders' messages. For example, while there was only one attested form across all games for expressing the number *1* (that is, a single *apple* token), there were 5 unique forms for expressing the number *2*, and 19 forms for expressing the number *12*. Furthermore, the share of trials exhibiting the most commonly used form for each number dropped as numbers increased. While all 62 occurrences of the number *1* were represented by a single apple, the most common expression for the number *2* was used 51 times, and the most common expression for *12* was used only 29 times. This is reflected in the Simpson's diversity index of messages used to represent sets of each cardinality, which is close to 0 when the same expression is always used to represent a cardinality, and close to 1 when many different expressions are used to represent a cardinality, with equal frequency. In our data, this index is closest to 0 for the smallest cardinalities, and is close to 1 for large cardinalities (Fig. 5C).

3.2.3. Rule use

We next performed several exploratory analyses to determine which strategies were employed. We first asked whether participants made use of different kinds of rules in their messages. To do this, we measured the Levenshtein distance[3] (Levenshtein, 1966) between messages that senders produced and each of four model strategies. For any given game, the distance of every message from a model system was summed over all trials, providing a measure of how closely the sender in that game adhered to the use of one strategy or another. Finally, the distance of each game to one or another strategy was compared, and each game was assigned the strategy of the model to which it had the shortest distance.

The four possible strategies we assessed were: 1-to-1, cumulative-additive, place-value, and a 'single token' strategy. The 1-to-1 strategy entails repeating one shape as many times as there were objects in the target set (e.g., using 4 apples to represent a set of 4). The cumulative-additive strategy involved summing tokens to communicate a number, but drawing on the more complex shapes to represent the numbers 2, 3, and 4 (e.g., using 2 cherry pairs plus 1 apple to represent 5). Under this strategy, expressing numbers greater than the largest image (the four leaf clover) involved repeating that image as many times as 4 could be divided into the number, with a remainder expressed via other shapes with smaller cardinal meanings. For example, the number 10 might be expressed as *clover-clover-cherries*, or *442*. The third strategy, which we expected to be quite unlikely, was a place-value system, similar to Arabic notation. In a system with only 4 possible shapes to use as symbols, the second place value represents multiples of 4, such that a shape representing 1 (e.g., apple) produces a value of 4, while a shape representing 2 (e.g., cherry) produces a value of 8, such that a string like *cherry-apple* represents 9. Finally, we modeled an ordinal system, where the characteristic cardinality of the shape chosen represented the ordinal position of the target among the distractors, ignoring their cardinality entirely (i.e., 1 apple for 1st set from the left, a pair of cherries for 2nd, etc.).

[3] A measure of string distance defined as the number of character insertions and deletions required to turn one string into another, often used in studying artificial symbol systems (e.g. Cornish, Dale, Kirby, & Christiansen, 2017; Vokey & Brooks, 1992).

This analysis suggested that most games relied on a cumulative-additive system (17 games), while a smaller number relied primarily on 1-to-1 correspondence (3 games). A cumulative additive system is also reflected in the frequency of each token (apple, cherries, oak leaves, and clover): the apple and clover were the most common tokens (Fig. 5B). Preliminary classifications suggested that 6 additional games were closest to the ordinal strategy, and 5 to the place-value system. However, on closer, qualitative inspection, the strategies in these latter two groups of games may reflect a naïve grouping strategy, where groups of objects in the target image are represented in order as they're viewed, so that 8 objects arranged into groups of 1, 2, 1, 3, and 1 are represented as 12,131. Given the observed distribution of strategies, the cumulative-additive strategy may be the only strategy reliably used by participants across our studies to feature a regular syntax – perhaps enabled by the change of communicative medium.

3.2.4. Order

Finally, one property of the cumulative-additive system that was common across games was the arrangement of tokens in a decreasing order, where 'smaller' tokens were always to the right of larger tokens (e.g., *442* is decreasingly ordered but *424* or *244* are not). Overall, 87.7% of messages were decreasingly ordered in this way, a number that did not significantly differ between correct and incorrect trials (88.1% and 86.3%). To understand whether this trend should be expected by chance, we generated a random message for each trial, such that the sum of all tokens in the participant's message and the random message were the same. We then used a permutation test to ask whether our random messages exhibited decreasing order as often as participants' messages, and found that participants' messages were significantly more likely to be decreasingly ordered ($t = 8.24$, $p < .001$). This preference was found across most games, and may have facilitated the process of reading those messages. Exploratory *t*-tests indicate that the reaction time of receivers was much faster when confronted with decreasingly ordered messages from their sender ($t = 5.21$, $p < .001$). More surprisingly, a similar trend holds for the time it took senders to construct their messages, as decreasingly ordered messages were also faster to make ($t = 5.742$, $p < .001$), though this trend may owe itself to faster reactions and better performance of those participants who also opted to use this strategy, rather than an effect of the strategy per se.

3.3. Discussion

In Experiment 3, we provided participants with pictures of familiar objects that had stable numerical properties (e.g., a pair of cherries, a four leaf clover), and asked whether they would exploit these properties to communicate about number. We found that few participants persisted in using 1-to-1, and that most instead used the features of provided objects to compress their representations of number. The prevalence of compressed strategies (84.6%) was slightly greater than in Experiment 2 (77.6%). Unlike Experiment 2, however, the modality of Experiment 3 required participants to arrange a set of discrete shapes in a linear order. Given this constraint, many participants invented a numerical base out of the 4-leaf clover shape with the result that most messages in shared two features that are characteristic of historically attested numeral systems. One feature was the use of a cumulative-additive structure, taking advantage of the highest available base (e.g., expressing 5 as *clover* + *apple*, rather than as *cherries* + *cherries* + *apple*). The other characteristic feature was the ordering of tokens from largest to smallest, similar not only to the familiar Arabic numeral system, and also to the majority of linear numeral systems in historical record (Chrisomalis, 2020).

4. General discussion

Across three experiments, we investigated how humans create and combine symbols to express number. In Experiment 1, we found that when participants were asked to communicate about number to a partner, they often created sketches that used a 1-to-1 correspondence strategy: For each object in a set, they generally created one corresponding dot, mark, or sketch of that thing. However, they sometimes used configural cues to express number, and in some cases, dyads used representations that could not be decoded by independent participants, potentially compatible with the creation of new conventions. In Experiment 2, we directly explored this finding, and found that when configural cues were reliably available across trials participants used 1-to-1 correspondence to communicate number much less frequently, and often used compressed representations that exploited configural cues. In Experiment 3, rather than requiring participants to invent new conventions de novo, we asked whether they would exploit publicly shared correlates of number by providing them with shapes that could act as candidate conventions (e.g., a pair of cherries, or a four-leaf clover). Here we found that participants rarely used 1-to-1 correspondence in favor of compressed representations, and also that when they combined shapes to communicate number, they often used a descending cumulative-additive structure, using the largest available "symbol" as a base (i.e., the clover).

These results suggest two main conclusions. First, the problem of creating entirely novel representations of number may be uniquely difficult when proxies for number, like shape, are not available in the communicative context. Whereas Shape game participants in Experiment 1 readily created conventions that used minimal strokes and were hard for naïve "recognizers" to decode (as in Garrod et al., 2007; Hawkins et al., 2023), participants in Number games invented conventions much less often, as reflected both by their persistent use of one-to-one correspondence and the relative ease with which recognizers identified the target numbers communicated by sketches. This was despite the fact that our participants were numerate adults who were familiar with western Arabic numerals. However, a second conclusion is that participants readily create compressed representations of number when non-numerical proxy representations are available in the context. Although most participants in Experiment 1 relied on 1-to-1 correspondence, some tried to communicate number by preserving the configural cues contained within stimuli. In Experiment 2, when these configural cues were more robust and reliable (i.e., arranging objects into rows in a consistent manner), participants frequently departed from the use of 1-to-1 strategies. Finally, this reliance on shape information was strongest when shapes of familiar objects - like apples, or clovers - could be combined to express number (Experiment 3).

As noted in the Introduction, humans often use ubiquitous and publicly available objects like hands, feet, and other body parts to count, and often use labels for these parts to subsequently name different cardinalities (Dahl, 1981; Epps, 2006; Heine, 2004; Rischel, 1997). Also, in at least some languages, small sets are labeled using names for common objects that feature specific numbers of things - like the number of seeds in a fruit, the number of eyes on one's head, or the number of fingers and toes on one's body (Epps, 2006). Similar practices of so-called ad hoc comparison can also be found in the history of measurement systems (Cooperrider & Gentner, 2019), but also in other cases such as color, where hues are labeled via metonymy, using the names of things in the local environment that have the same color (e.g., orange, lilac, gold, etc.; Casson, 1994). The logic underlying such instances of metonymy may help explain why participants in Experiments 1 and 2 did not create novel conventions as readily as in other studies of graphical communication. For example, in a study by Fay et al. (2010) some participants conveyed the concept *parliament* by initially sketching members of parliament sitting at tables with the Australian flag, but later included only the flag, which sufficed as an index of the concept even without the depiction of the people. This was possible because the concept parliament is associated with multiple imageable components which, after being used collectively to identify the intended referent, can be subsequently reduced to only one element (e.g., a flag) to create a convention. However, such a strategy is often not available in the case of number,

since in a set of, e.g., four ducks, no individual duck has any feature that can alone communicate the cardinality four. The use of 1-to-1 correspondence appears to be overcome only when there exists an imageable correlate of number, such as a spatial configuration, or a shape or name that is strongly associated with a particular cardinality (e.g., eyes, clovers, etc.).

In addition to finding that participants often use physical correlates of number to communicate about large quantities, the results from this study also suggest ways in which the medium of numerical symbols may impact the form that they take. Whereas the reference sets in all experiments – as well as the communicative medium in Experiments 2 and 3 – were two-dimensional visual arrays, only in Experiment 3 did the communicative medium require a strict linear ordering of component shapes. This appears to have led participants to create combinatorial representations of number, perhaps by prompting them to reflect on their strategy for encoding the objects in the target set. This is potentially important, because previous theories of the history of number posit that abstraction occurs when representations transition from purely referential meanings towards meanings that are defined by relations between elements within the symbolic system. New abstractions, like the rules that govern written numerals, are made possible by the affordances of the symbolic medium in which the system is instantiated - in this case a linearly ordered sequence (Overmann, 2018).

While this collection of exploratory studies provides a first step towards understanding the processes that underpin the creation of numeral systems, they had several notable limitations. One is that some of our pre-registered analyses proved intractable in the face of the data we collected. For example, a planned analysis of spatial clustering of strokes within sketches was unable to meaningfully recover coherent clusters that were apparent to the human eye (e.g., dotted lines drawn by participants), rendering results of that analysis questionable. Also, we initially planned to conduct a recognition task in Experiment 2 that paralleled Experiment 1, but found that participants gave highly divergent and often uninterpretable responses to sketches, making it unclear whether the labeling techniques we offered them were understood. Another limitation is that our methods did not make it possible to easily identify the cognitive steps involved in the creation of number representations. For example, it is possible that some participants in Experiments 1 and 2 did not even consider creating summary representations of cardinality, and instead tried to communicate the visual patterns present in arrays, rather than number.

Another limitation, common to communicative games, is that participants in our study could only communicate via the medium we provided to them, and were not allowed to directly tell partners what symbols meant. This constraint was important to precluding uninteresting strategies in our numerate participants (like creating a direct translation of existing numerals). If laboratory participants were allowed to make explicit agreements, they could easily bypass the most important communicative obstacle to creating numerals from scratch by using known symbols as anchors for creating new ones - e.g., by explicitly agreeing to replace “5” with “%” (or some novel form). Also, we felt that our method might simulate the challenge of expressing exact number when existing symbols are absent, since this requires expressing number via some other medium or strategy. Still, most number systems almost certainly evolved through processes of *conspiracy* among community members, in which individuals worked together to create, agree upon, and teach new symbols. Exploring how such collaborative processes might work when existing numeral conventions are not available should be explored further in future work. Indeed, another limitation of our study is that all of our participants were numerate adults. Although their numeracy proved to be of little help in creating novel systems, it remains an interesting question whether innumerate adults or young children might deploy different strategies - a question we are currently exploring.

In summary, results from three experiments suggest that when numerate adults are tasked with devising novel ways to communicate about number, they often default to using 1-to-1 correspondence, but also readily exploit visual correlates of number to efficiently convey representations (e.g., configural cues or objects that canonically appear in certain numbers). Features of the medium of communication may also influence the creation of conventions by limiting degrees of freedom, and by prompting communicators to spontaneously organize their messages in novel ways.

Funding

Funding sources were: NSF Award #2000827 to DB; NSF CAREER #2047191 to JEF.

Data repository

All pre-registrations, materials, data, and analyses are available at the authors’ Github project repository: https://github.com/cogtoolslab/improvised_numerals. Pre-registrations can also be found on the OSF repository: https://osf.io/4q3t9.

CRediT authorship contribution statement

Sebastian Holt: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Judith E. Fan:** Conceptualization, Methodology, Software, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **David Barner:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors have no interests that may influence or appear to have influenced the above work. AI was not used in the writing process.

Data availability

https://github.com/cogtoolslab/improvised_numerals.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105665.

References

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: A new look at visual numerosity judgements. *Perception, 5*(3), 327–334.
Barner, D. (2017). Language, procedures, and the non-perceptual origin of number word meanings. *Journal of Child Language, 44*(3), 553–590.
Beller, S., Bender, A., Chrisomalis, S., Jordan, F. M., Overmann, K. A., Saxe, G. B., & Schlimm, D. (2018). The cultural challenge in mathematical cognition. *Journal of Numerical Cognition, 4*(2).
Bender, A., & Beller, S. (2011). Fingers as a tool for counting–naturally fixed or culturally flexible? *Frontiers in Psychology, 2*, 256.
Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition, 124*(2), 156–182.
Brannon, E. M., & Merritt, D. J. (2011). Evolutionary foundations of the approximate number system. *Space, Time and Number in the Brain*, 207–224.
Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59–68.
Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences, 23*(10), 823–835.
Casson, R. W. (1994). Russett, rose, and raspberry: The development of English secondary color terms. *Journal of Linguistic Anthropology, 4*(1), 5–22.
Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences, 272*(1560), 267–275.
Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Cambridge Archaeological Journal, 14*(1), 37–52.
Chrisomalis, S. (2010). *Numerical notation: A comparative history*. Cambridge University Press.

Chrisomalis, S. (2020). *Reckonings: Numerals, cognition, and history*. MIT Press.
Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300–307.
Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind, 4*, 102–118.
Comrie, B. (2011). Typology of numeral systems. Numeral types and changes worldwide. *Trends in Linguistics. Studies and Monographs*, 118.
Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition, 191*, Article 103942.
Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PLoS One, 12*(1), Article e0168532.
Dahl, O. C. (1981). Austronesian numerals. In R. A. Blust (Ed.), *Historical linguistics in Indonesia, part I* (pp. 46–58). Jakarta: Universitas Atma Jaya.
Damerow, P. (1996). Number as a second-order concept. *Science in Context, 9*(2), 139–149.
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.
Epps, P. (2006). Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica, 23*(2), 259–288.
Everett, C., & Madora, K. (2012). Quantity recognition among speakers of an anumeric language. *Cognitive Science, 36*(1), 130–141.
Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34*(3), 351–386.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.
Frank, M., Everett, D., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*(3), 819–824.
Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*(1), 134.
Gallistel, C. R., Gelman, R., & Cordes, S. (2006). 12 the cultural and evolutionary history of the real numbers. *Evolution and Culture, 247*.
Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science, 31*(6), 961–987.
Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389–407.
Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499.
Hammarström, H. (2010). Rarities in numeral systems. In J. Wohlgemuth, & M. Cysouw (Eds.), *Rethinking universals: How rarities affect linguistic theory* (pp. 11–60). Berlin: Mouton De Gruyter.
Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition, 15*(1–3), 95–110.
Hawkins, R. D., Sano, M., Goodman, N. D., & Fan, J. E. (2023). Visual resemblance and interaction history jointly constrain pictorial meaning. *Nature Communications, 14*(1), 2199.
Heine, B. (2004). On genetic motivation in grammar. In G. Radden, & K. U. Panther (Eds.), *Studies in linguistic motivation (pp. 103–120)*. Berlin/New York: Mouton de Gruyter.
Hofer, M., & Levy, R. P. (2019). Iconicity and structure in the emergence of combinatoriality. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.
Høyrup, J. (1994). *In measure, number, and weight: Studies in mathematics and culture*. SUNY Press.
Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system*. Oxford: Basil Blackwell.
Ifrah, G. (2000). *The universal history of numbers*. London: Harvill.
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681–10686.
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics – Doklady, 10*(8), 707–710.
Malafouris, L. (2010). *Grasping the concept of number: How did the sapient mind move beyond approximation. The archaeology of measurement: Comprehending heaven, earth and time in ancient societies* (pp. 35–42).
Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1.
O'Shaughnessy, D. M., Gibson, E., & Piantadosi, S. T. (2021). The cultural origins of symbolic number. *Psychological Review, 129*(6), 1442–1456.
Overmann, K. A. (2018). Constructing a concept of number. *Journal of Numerical Cognition, 4*(2), 464–493.
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499–503.
Rischel, J. (1997). Typology and reconstruction of numeral systems: The case of Austroasiatic. In J. Fisiak (Ed.), *Linguistic reconstruction and typology* (pp. 273–312). Berlin/New York: Mouton de Gruyter.
Robson, E. (2008). *Mathematics in ancient Iraq: A social history*. Princeton University Press.
Saxe, G. B. (1981). Body parts as numerals: A developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development, 52*(1), 306–316.
Saxe, G. B., & Esmonde, I. (2012). *Cultural development of mathematical ideas: Papua New Guinea studies*. Cambridge University Press.
Schmandt-Besserat, D. (1978). The earliest precursor of writing. *Scientific American, 238*(6), 50–59.
Schmandt-Besserat, D. (2010). *How writing came about*. University of Texas Press.
Schneider, R. M., Brockbank, E., Feiman, R., & Barner, D. (2022). Counting and the ontogenetic origins of exact equality. *Cognition, 218*, Article 104952.
Srinivasan, M., Wagner, K., Frank, M. C., & Barner, D. (2018). The role of design and training in artifact expertise: The case of the abacus and visual attention. *Cognitive Science, 42*, 757–782.
Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120–137.
Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology, 16*(2), 145–176.
Van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics, 32*, 163–170.
Verhoef, T., Kirby, S., & De Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science, 40*(8), 1969–1994.
Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 328.
Wang, Y., Lew, T. F., Brady, T. F., & Vul, E. (2023). Structured visuospatial representations revealed through serial reproduction. *Journal of Experimental Psychology. Human Perception and Performance, 49*(6), 862–876.
Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137.
Williams, F. E. (1940). Natives of Lake Kutubu, Papua. *Oceania, 11*(2), 121–157. https://doi.org/10.1002/j.1834-4461.1940.tb00281.x
Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind, 4*, 57–70.